Final Project

Created	@December 10, 2024
Due Date	@December 15, 2024 11:59 PM
Tags	Course 7

Introduction:

Life expectancy, defined as the average number of years a person can expect to live from birth, is a crucial indicator of well-being and development. While traditionally measured on a scale from 0 to 100 years, this metric encompasses various complex factors that influence human longevity. In an era where global health challenges are becoming increasingly prominent, understanding the determinants of life expectancy has never been more critical. This analysis aims to explore the intricate relationship between life expectancy and various socioeconomic factors across different nations. While preliminary research suggests a disparity between developed and developing countries, primarily attributed to resource availability, the reality may be more nuanced. My investigation will delve deeper into whether this developed-developing nation divide is as straightforward as it appears, or if other significant factors play equally important roles in determining life expectancy. Furthermore, we will examine the potential correlation between life expectancy and lifestyle factors, considering how these relationships might vary across different geographical and economic contexts. This comprehensive analysis seeks to uncover the multifaceted nature of factors influencing human longevity and challenge conventional assumptions about the determinants of life expectancy.

Methodology:

The methodology for this analysis consisted of three main phases. First, the data processing and cleaning phase involved handling empty values and performing necessary dataset cleaning to ensure data quality and reliability. The second phase focused on Exploratory Data Analysis (EDA), where I conducted an initial exploration to understand the distributions and relationships within the data. This included calculating basic statistical measures and generating visualizations to identify and illustrate key trends. In the final phase, I employed graph algorithms to model relationships between

countries based on their life expectancy data. This involved creating a network where countries were connected based on similar health outcomes, followed by graph clustering and partitioning to identify representative nodes. The effectiveness of this clustering was then evaluated against prior knowledge and expectations about the network structure.

Results:

Statistics:

Plot of Income vs Schooling rates across different countries and years

Analysis of the data reveals a strong positive correlation between income levels and education across countries. The visualization demonstrates that nations with higher average incomes consistently exhibit higher levels of education among their populations. This relationship suggests that educational achievement is a significant predictor of economic growth, as advanced education often leads to increased employment opportunities and higher earning potential.

Top 5 Countries with high life expectancies across each year (2000 - 2015)

```
Top 5 countries in year 2000:
Sierra Leone: 39.00
Malawi: 43.10
Zambia: 43.80
Angola: 45.30
Eritrea: 45.30
```

Year 2000

```
Top 5 countries in year 2001:
Sierra Leone: 41.00
Malawi: 43.50
Zambia: 44.60
Zimbabwe: 45.30
Central African Republic: 45.60
```

Year 2001

```
Top 5 countries in year 2002:
Malawi: 44.00
Zimbabwe: 44.80
Zambia: 45.50
Central African Republic: 45.60
Botswana: 46.00
```

Year 2002

```
Top 5 countries in year 2004:
Sierra Leone: 42.30
Zimbabwe: 44.30
Lesotho: 44.80
Malawi: 45.10
Swaziland: 45.60
```

Year 2004

```
Top 5 countries in year 2003:
Sierra Leone: 41.50
Zimbabwe: 44.50
Malawi: 44.60
Lesotho: 45.50
Central African Republic: 45.70
```

Year 2003

```
Top 5 countries in year 2005:
Sierra Leone: 43.30
Lesotho: 44.50
Zimbabwe: 44.60
Central African Republic: 45.90
Malawi: 46.00
```

Year 2005

```
Top 5 countries in year 2006:
Sierra Leone: 44.30
Lesotho: 45.30
Zimbabwe: 45.40
Central African Republic: 46.30
Malawi: 47.10
```

Year 2006

```
Top 5 countries in year 2007:
Sierra Leone: 45.30
Lesotho: 46.20
Zimbabwe: 46.60
Central African Republic: 46.80
Angola: 48.20
```

Year 2007

```
Top 5 countries in year 2008:
Sierra Leone: 46.20
Central African Republic: 47.60
Lesotho: 47.80
Zimbabwe: 48.20
Angola: 48.70
```

Year 2008

```
Top 5 countries in year 2009:
Sierra Leone: 47.10
Central African Republic: 48.60
Angola: 49.10
Lesotho: 49.40
Zimbabwe: 50.00
```

Year 2009

```
Top 5 countries in year 2010:
Haiti: 36.30
Sierra Leone: 48.10
Central African Republic: 49.20
Angola: 49.60
Lesotho: 51.10
```

Year 2010

```
Top 5 countries in year 2011:
Sierra Leone: 48.90
Central African Republic: 49.80
Angola: 51.00
Chad: 51.60
Côte d'Ivoire: 51.70
```

Year 2011

```
Top 5 countries in year 2012:
Sierra Leone: 49.70
Chad: 51.80
Côte d'Ivoire: 52.00
Lesotho: 52.20
Nigeria: 52.70
```

Year 2012

```
Top 5 countries in year 2013:
Cook Islands: 0.00
Dominica: 0.00
Marshall Islands: 0.00
Monaco: 0.00
Nauru: 0.00
```

Year 2013

```
Top 5 countries in year 2014:
Sierra Leone: 48.10
Angola: 51.70
Lesotho: 52.10
Chad: 52.60
Côte d'Ivoire: 52.80
```

Year 2014

```
Top 5 countries in year 2015:
Sierra Leone: 51.00
Angola: 52.40
Central African Republic: 52.50
Chad: 53.10
Côte d'Ivoire: 53.30
```

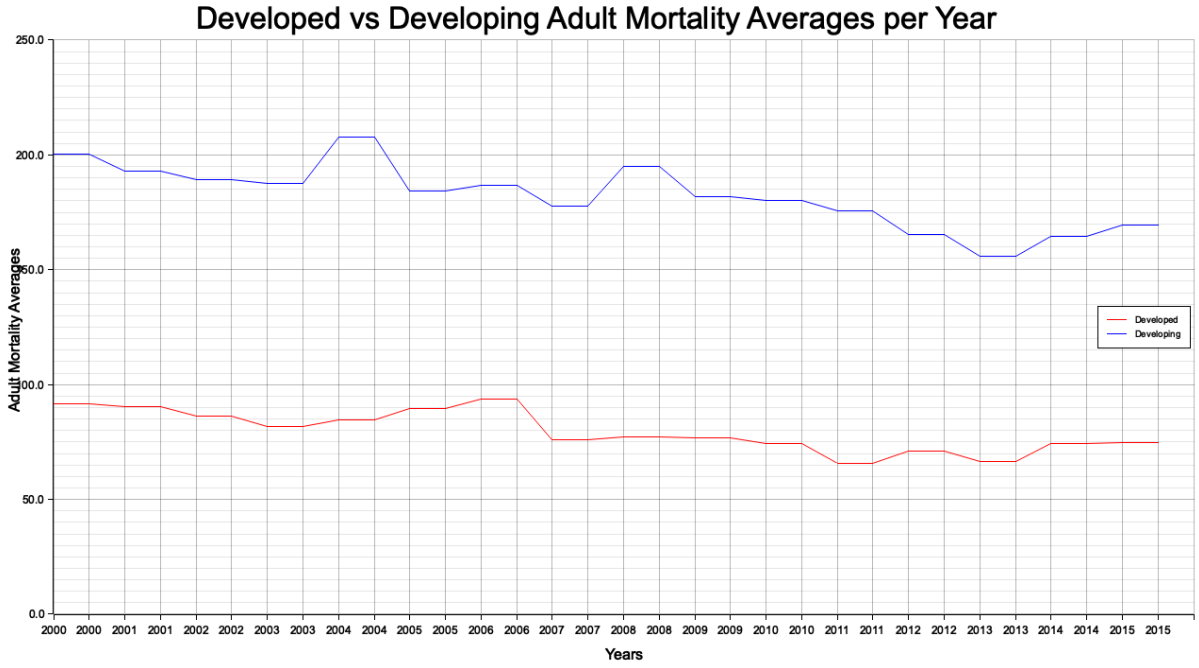
Year 2015

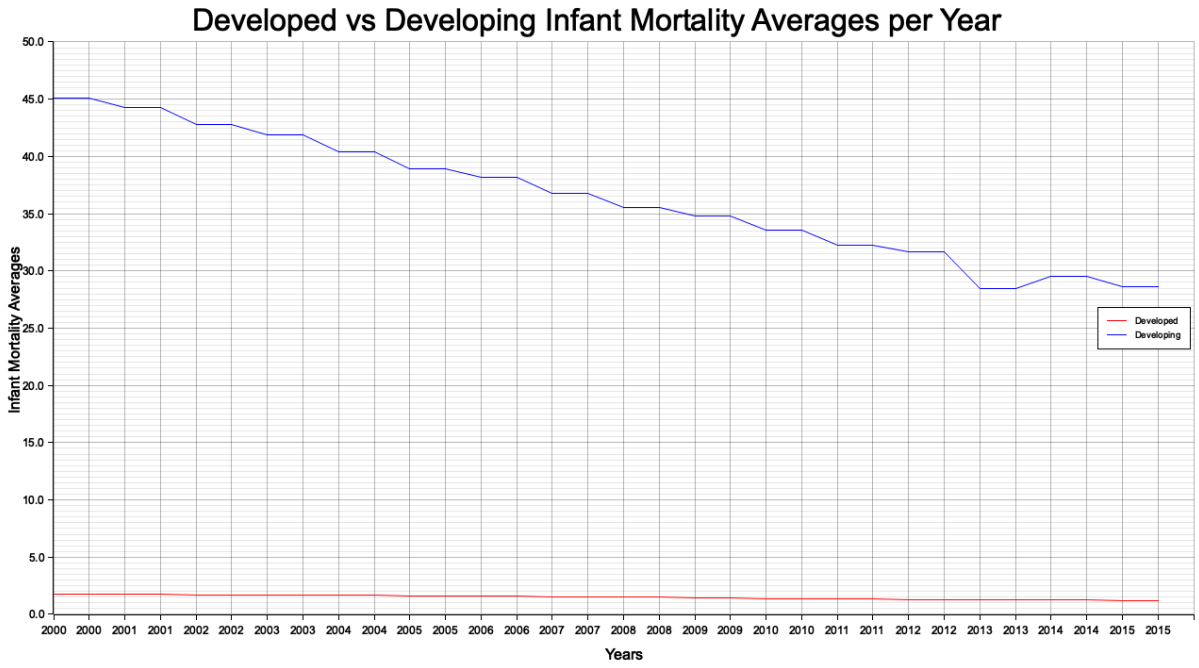
Analysis of life expectancy data from 2000-2015 revealed surprising findings. The top 5 countries consistently maintaining high life expectancies were Sierra Leone, Central African Republic, Angola, Lesotho, and Zimbabwe. This finding challenges conventional ideas, as these nations are typically categorized as developing countries. While the National Institute of Health (NIH) research traditionally indicates lower life expectancies in developing nations, my analysis demonstrates that these countries actually maintained higher life expectancy rates compared to their developed counterparts during this period. Sierra Leone, in particular, showed remarkable sustained high life expectancy levels throughout the 15 years.

The average life expectancy across developed countries stands at 79 years, while developing countries average 67 years.

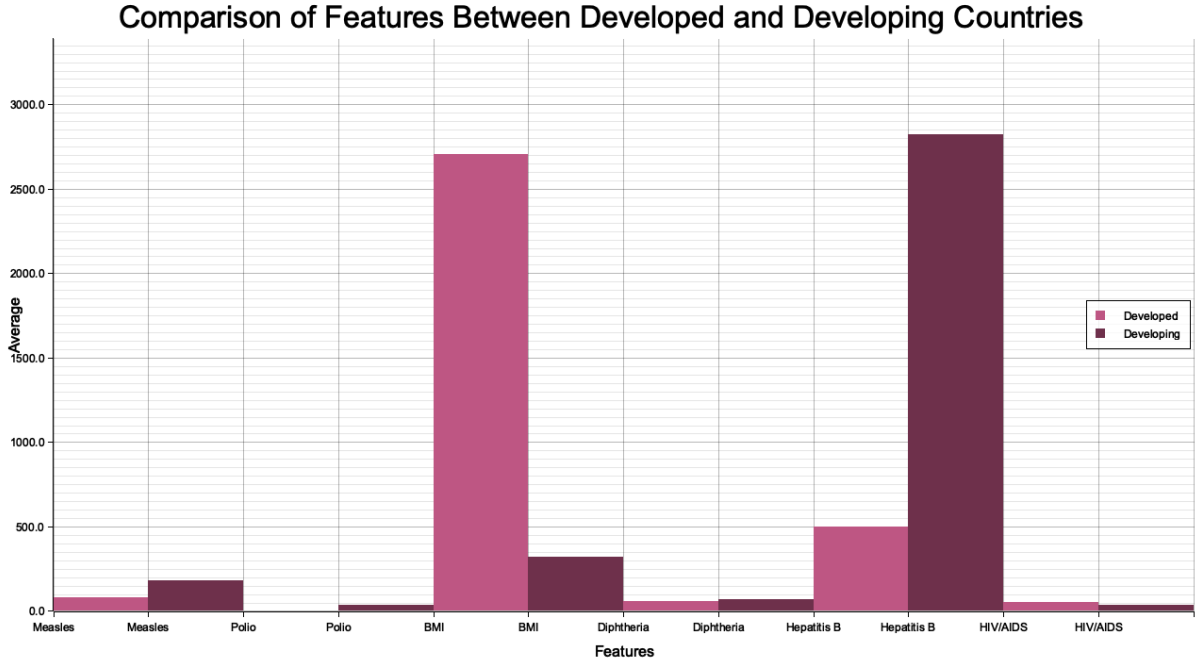
This data aligns with established research findings, confirming that developed nations generally maintain higher life expectancy rates when examined on a larger scale.

Developed vs Developing: Adult Mortality and Infant Death Averages





Further analysis revealed distinct mortality patterns between developing and developed nations. Developing countries exhibited higher adult and infant mortality rates, though they showed consistent improvement in infant mortality over the 15-year period. In contrast, developed countries maintained stable adult mortality rates and demonstrated consistently low infant mortality rates with minimal fluctuation.



Developing countries showed higher levels of diseases including measles, polio, diphtheria, and hepatitis B. These higher disease rates, combined with limited access to healthcare and treatment resources, it is safe to say, contribute to increased mortality rates in these regions. In contrast, developed nations face different health challenges - notably higher rates of obesity (as measured by BMI) and HIV/AIDS. However, their healthcare systems and better access to medical treatments help maintain lower overall mortality rates despite these health issues.

Graph Algorithm:

The graph algorithm used in this analysis utilized clustering based on three key features: life expectancy, GDP, and population. This clustering approach revealed distinct groupings of countries with similar characteristics across these dimensions. The resulting clusters effectively identified nations with comparable socio-economic profiles, providing insights into the patterns and relationships between these fundamental development indicators.

Niue (Cluster 1909): represents a group of small, isolated countries with similar socio-economic characteristics (e.g., low population, specific GDP, and life expectancy values).

Albania (Cluster 0): represents countries with medium development metrics, often transitional economies with improving life expectancy and GDP.

Cook Islands (Cluster 624): represents small island nations with specific population and economic structures, similar to Niue.

Monaco (Cluster 1715): represents wealthy, small-population countries with extremely high GDP and life expectancy.

Saint Kitts and Nevis / San Marino (Clusters 2167, 2216): represent micro-states or small, affluent countries that share unique economic and demographic profiles.

Conclusion:

My analysis revealed a clear divide between developed and developing countries in terms of life expectancy. Developed countries show an average lifespan of 79 years, while developing nations average 67 years. This gap can be largely attributed to differences in healthcare systems and disease patterns. Developing countries struggle with diseases like measles, polio, diphtheria, and hepatitis B, compounded by limited access to healthcare. While developed nations face their own health challenges, such as higher rates of obesity and HIV/AIDS, their superior healthcare systems help maintain lower mortality rates. I also observed that developing countries are making progress, particularly in reducing infant mortality rates, though their overall adult and infant mortality rates remain higher than those of developed nations. Additionally, my research found that education plays a crucial role in this disparity - countries with higher education levels tend to have stronger economies and better health outcomes. When we grouped countries based on life expectancy, GDP, and population, we found diverse clusters ranging from small, isolated nations to wealthy micro-states, highlighting how these factors work together to influence overall life expectancy.

Ai Usage:

I used ChatGPT to assist with various aspects of my work, from function creation to visualizations. Initially, I was able to read and clean my data independently by referencing class examples. However, I encountered difficulties when handling missing values. My first approach was to fill them with zeros, but this led to inaccurate data. ChatGPT suggested using NaN to fill

missing values, which preserved the integrity of my dataset. Furthermore, ChatGPT assisted in creating a correlation heat map function. Although I struggled to label each feature's correlation, with its help, I was able to create a heat map that visually represented correlations. This visualization involved parsing features through a vector to chart individual correlations and push non-correlations. Additionally, ChatGPT helped me compute correlations using the Pearson correlation coefficient between two input arrays, x and y. It also helped me in implementing a graph algorithm to cluster countries with similar features, which was then exported into a CSV file. This experience taught me how to write CSV files similarly to Python's 'w' mode. Moreover, I used ChatGPT in creating a bar chart showing disease statistics averaged by development status (developing and developed). My initial attempt resulted in a stacked bar graph, but ChatGPT helped me create a multi-bar graph that displayed both statistics side by side. Additionally, ChatGPT was important in resolving any code errors I encountered. It suggested alternatives for deprecated code and provided solutions for running my code in my IDE. Overall, ChatGPT was instrumental in helping me navigate through challenges and enhance my data analysis, from data cleaning to complex visualizations.

Citations:

Rogers, R G, and S Wofford. “Life expectancy in less developed countries: socioeconomic development or public health?.” *Journal of biosocial science* vol. 21,2 (1989): 245-52. doi:10.1017/s0021932000017934

Dataset : Life Expectancy (WHO)